

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Matalon, Reuben
Kaul, Rajinder
Gao, Guang Ping
Balamurugan, Kuppareddi
Michals-Matalon, Kimberlee

(ii) TITLE OF INVENTION: Aspartoacylase Gene, Protein, and
Methods of Screening for Mutations Associated with Canavan
Disease

(iii) NUMBER OF SEQUENCES: 27

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
(B) STREET: 2200 Clarendon Boulevard, Suite 1400
(C) CITY: Arlington
(D) STATE: Virginia
(E) COUNTRY: U.S.A.
(F) ZIP: 22201

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/128,020
(B) FILING DATE: 29-SEP-1993
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Hamlet-King, Diana
(B) REGISTRATION NUMBER: 33,302
(C) REFERENCE/DOCKET NUMBER: Shutt 1

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 703-243-6333
(B) TELEFAX: 703-243-6410
(C) TELEX: 64191

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1435 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 159..1097

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TTGTAACAGA AAATTTAAAT ATACTCCACT CAAGGGAATT CTGTACTTTG CCCTTTGGT

60

AAAGTCTCAT TTACATTTCT AACACTTTCT TAAGAAAATC GAATTCCTT TGATCTCTCT

120

TCTGAATTGC AGAAATCAGA TAAAAACTAC TTGGTGAA ATG ACT TCT TGT CAC Met Thr Ser Cys His 1 5	173
ATT GCT GAA GAA CAT ATA CAA AAG GTT GCT ATC TTT GGA GGA ACC CAT Ile Ala Glu Glu His Ile Gln Lys Val Ala Ile Phe Gly Gly Thr His 10 15 20	221
GGG AAT GAG CTA ACC GGA GTA TTT CTG GTT AAG CAT TGG CTA GAG AAT Gly Asn Glu Leu Thr Gly Val Phe Leu Val Lys His Trp Leu Glu Asn 25 30 35	269
GGC GCT GAG ATT CAG AGA ACA GGG CTG GAG GTA AAA CCA TTT ATT ACT Gly Ala Glu Ile Gln Arg Thr Gly Leu Glu Val Lys Pro Phe Ile Thr 40 45 50	317
AAC CCC AGA GCA GTG AAG AAG TGT ACC AGA TAT ATT GAC TGT GAC CTG Asn Pro Arg Ala Val Lys Lys Cys Thr Arg Tyr Ile Asp Cys Asp Leu 55 60 65	365
AAT CGC ATT TTT GAC CTT GAA AAT CTT GGC AAA AAA ATG TCA GAA GAT Asn Arg Ile Phe Asp Leu Glu Asn Leu Gly Lys Lys Met Ser Glu Asp 70 75 80 85	413
TTG CCA TAT GAA GTG AGA AGG GCT CAA GAA ATA AAT CAT TTA TTT GGT Leu Pro Tyr Glu Val Arg Arg Ala Gln Glu Ile Asn His Leu Phe Gly 90 95 100	461
CCA AAA GAC AGT GAA GAT TCC TAT GAC ATT ATT TTT GAC CTT CAC AAC Pro Lys Asp Ser Glu Asp Ser Tyr Asp Ile Ile Phe Asp Leu His Asn 105 110 115	509
ACC ACC TCT AAC ATG GGG TGC ACT CTT ATT CTT GAG GAT TCC AGG AAT Thr Thr Ser Asn Met Gly Cys Thr Leu Ile Leu Glu Asp Ser Arg Asn 120 125 130	557
AAC TTT TTA ATT CAG ATG TTT CAT TAC ATT AAG ACT TCT CTG GCT CCA Asn Phe Leu Ile Gln Met Phe His Tyr Ile Lys Thr Ser Leu Ala Pro 135 140 145	605
CTA CCC TGC TAC GTT TAT CTG ATT GAG CAT CCT TCC CTC AAA TAT GCG Leu Pro Cys Tyr Val Tyr Leu Ile Glu His Pro Ser Leu Lys Tyr Ala 150 155 160 165	653
ACC ACT CGT TCC ATA GCC AAG TAT CCT GTG GGT ATA GAA GTT GGT CCT Thr Thr Arg Ser Ile Ala Lys Tyr Pro Val Gly Ile Glu Val Gly Pro 170 175 180	701
CAG CCT CAA GGG GTT CTG AGA GCT GAT ATC TTG GAT CAA ATG AGA AAA Gln Pro Gln Gly Val Leu Arg Ala Asp Ile Leu Asp Gln Met Arg Lys 185 190 195	749
ATG ATT AAA CAT GCT CTT GAT TTT ATA CAT CAT TTC AAT GAA GGA AAA Met Ile Lys His Ala Leu Asp Phe Ile His His Phe Asn Glu Gly Lys 200 205 210	797
GAA TTT CCT CCC TGC GCC ATT GAG GTC TAT AAA ATT ATA GAG AAA GTT Glu Phe Pro Pro Cys Ala Ile Glu Val Tyr Lys Ile Ile Glu Lys Val 215 220 225	845

25iii

GAT TAC CCC CGG GAT GAA AAT GGA GAA ATT GCT GCT ATC ATC CAT CCT Asp Tyr Pro Arg Asp Glu Asn Gly Glu Ile Ala Ala Ile Ile His Pro 230 235 240 245	893
AAT CTG CAG GAT CAA GAC TGG AAA CCA CTG CAT CCT GGG GAT CCC ATG Asn Leu Gln Asp Gln Asp Trp Lys Pro Leu His Pro Gly Asp Pro Met 250 255 260	941
TTT TTA ACT CTT GAT GGG AAG ACG ATC CCA CTG GGC GGA GAC TGT ACC Phe Leu Thr Leu Asp Gly Lys Thr Ile Pro Leu Gly Gly Asp Cys Thr 265 270 275	989
GTG TAC CCC GTG TTT GTG AAT GAG GCC GCA TAT TAC GAA AAG AAA GAA Val Tyr Pro Val Phe Val Asn Glu Ala Ala Tyr Tyr Glu Lys Lys Glu 280 285 290	1037
GCT TTT GCA AAG AGA ACT AAA CTA ACG CTC AAT GCA AAA AGT ATT CGC Ala Phe Ala Lys Thr Thr Lys Leu Thr Leu Asn Ala Lys Ser Ile Arg 295 300 305	1085
TGC TGT TTA CAT TAGAAATCAC TTCCAGCTTA CATCTTACAC GGTGTCTTAC Cys Cys Leu His 310	1137
AAATTCTGCTAGCTCTGTAAG CTCTTAAGA GTAGGGTTGT GCCTTATTCA ACTGCATAACA	1197
TAGCTCCTAG CACAGTGCCT TATTGGTAG GCATCTAACG AAATTTCTTA ATTAAATTAA	1257
TATATCTTTA AAGATATCAT ATTTATGTA TGTAGCTTAT TCAAAGAAGT GTTTCCTATT	1317
TCTATATAGT TTATTATACA TGATACTTGG GTAGCTAAC ATTCTTAATA AACAGCCTTT	1377
GTATTCTAGAA TATAAAATTG AAA TAGATAT ATATAAGTT AAAAAAAA AAAAAAAA	1435

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 313 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 83
- (D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 105
- (D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 108
- (D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 146
- (D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site,

(B) LOCATION: 264
 (D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

(A) NAME/KEY: Modified-site
 (B) LOCATION: 117
 (D) OTHER INFORMATION: /note= "Potential N-glycosylation site"

(ix) FEATURE:

(A) NAME/KEY: Active-site
 (B) LOCATION: 18..24
 (D) OTHER INFORMATION: /note= "Consensus sequence predicted to be involved in catalysis"

(ix) FEATURE:

(A) NAME/KEY: Active-site
 (B) LOCATION: 275..278
 (D) OTHER INFORMATION: /note= "Consensus sequence predicted to be involved in catalysis"

(ix) FEATURE:

(A) NAME/KEY: Active-site
 (B) LOCATION: 283..289
 (D) OTHER INFORMATION: /note= "Consensus sequence predicted to be involved in catalysis"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Thr Ser Cys His Ile Ala Glu Glu His Ile Gln Val Ala Ile
 1 5 10 15

Phe Gly Gly Thr His Gly Asn Glu Leu Thr Gly Val Phe Leu Val Lys
 20 25 30

His Trp Leu Glu Asn Gly Ala Glu Ile Gln Arg Thr Gly Leu Glu Val
 35 40 45

Lys Pro Phe Ile Thr Asn Pro Arg Ala Val Lys Lys Cys Thr Arg Tyr
 50 55 60

Ile Asp Cys Asp Leu Asn Arg Ile Phe Asp Leu Glu Asn Leu Gly Lys
 65 70 75 80

Lys Met Ser Glu Asp Leu Pro Tyr Glu Val Arg Arg Ala Gln Glu Ile
 85 90 95

Asn His Leu Phe Gly Pro Lys Asp Ser Glu Asp Ser Tyr Asp Ile Ile
 100 105 110

Phe Asp Leu His Asn Thr Thr Ser Asn Met Gly Cys Thr Leu Ile Leu
 115 120 125

Glu Asp Ser Arg Asn Asn Phe Leu Ile Gln Met Phe His Tyr Ile Lys
 130 135 140

Thr Ser Leu Ala Pro Leu Pro Cys Tyr Val Tyr Leu Ile Glu His Pro
 145 150 155 160

Ser Leu Lys Tyr Ala Thr Thr Arg Ser Ile Ala Lys Tyr Pro Val Gly
 165 170 175

Ile Glu Val Gly Pro Gln Pro Gln Gly Val Leu Arg Ala Asp Ile Leu
 180 185 190

Asp Gln Met Arg Lys Met Ile Lys His Ala Leu Asp Phe Ile His His
 195 200 205
 Phe Asn Glu Gly Lys Glu Phe Pro Pro Cys Ala Ile Glu Val Tyr Lys
 210 215 220
 Ile Ile Glu Lys Val Asp Tyr Pro Arg Asp Glu Asn Gly Glu Ile Ala
 225 230 235 240
 Ala Ile Ile His Pro Asn Leu Gln Asp Gln Asp Trp Lys Pro Leu His
 245 250 255
 Pro Gly Asp Pro Met Phe Leu Thr Leu Asp Gly Lys Thr Ile Pro Leu
 260 265 270
 Gly Gly Asp Cys Thr Val Tyr Pro Val Phe Val Asn Glu Ala Ala Tyr
 275 280 285
 Tyr Glu Lys Lys Glu Ala Phe Ala Lys Thr Thr Lys Leu Thr Leu Asn
 290 295 300
 Ala Lys Ser Ile Arg Cys Cys Leu His
 305 310

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 313 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /note= "This is isoleucine in human, valine in bovine. This is a very conservative substitution."

(ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /note= "This is glutamic acid in human, aspartic acid in bovine. This is a very conservative substitution."

(ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /note= "This is histidine in human, proline in bovine. This is a conservative substitution."

(ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: 12
- (D) OTHER INFORMATION: /note= "This is glutamine in human, lysine in bovine. This is a very conservative substitution."

(ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: 38

- (D) OTHER INFORMATION: /note= "This is glycine in human, serine in bovine. This is a very conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 39
(D) OTHER INFORMATION: /note= "This is alanine in human, threonine in bovine. This is a very conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 72
(D) OTHER INFORMATION: /note= "This is isoleucine in human, valine in bovine. This is a very conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 75
(D) OTHER INFORMATION: /note= "This is leucine in human, proline in bovine. This is not a conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 82
(D) OTHER INFORMATION: /note= "This is methionine in human, lysine in bovine. This is a conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 134
(D) OTHER INFORMATION: /note= "This is asparagine in human, aspartic acid in bovine. This is a very conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 200
(D) OTHER INFORMATION: /note= "This is lysine in human, glutamine in bovine. This is a very conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 208
(D) OTHER INFORMATION: /note= "This is histidine in human, asparagine in bovine. This is a very conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 226
(D) OTHER INFORMATION: /note= "This is isoleucine in human, methionine in bovine. This is a very conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 227

- (D) OTHER INFORMATION: /note= "This is glutamic acid in human, arginine in bovine. This is not a conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 234
(D) OTHER INFORMATION: /note= "This is aspartic acid in human, asparagine in bovine. This is a very conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 236
(D) OTHER INFORMATION: /note= "This is asparagine in human, serine in bovine. This is a very conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 240
(D) OTHER INFORMATION: /note= "This is alanine in human, serine in bovine. This is a very conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 246
(D) OTHER INFORMATION: /note= "This is asparagine in human, lysine in bovine. This is a very conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 258
(D) OTHER INFORMATION: /note= "This is glycine in human, glutamic acid in bovine. This is a conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 261
(D) OTHER INFORMATION: /note= "This is methionine in human, valine in bovine. This is a very conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 276
(D) OTHER INFORMATION: /note= "This is cysteine in human, glutamine in bovine. This is not a conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 306
(D) OTHER INFORMATION: /note= "This is lysine in human, asparagine in bovine. This is a very conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 310

- (D) OTHER INFORMATION: /note= "This is cysteine in human.
serine in bovine. This is a conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 311
(D) OTHER INFORMATION: /note= "This is cysteine in human.
serine in bovine. This is a conservative substitution."
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 83
(D) OTHER INFORMATION: /note= "Phosphorylation site"
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 105
(D) OTHER INFORMATION: /note= "Phosphorylation site"
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 108
(D) OTHER INFORMATION: /note= "Phosphorylation site"
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 146
(D) OTHER INFORMATION: /note= "Phosphorylation site"
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 264
(D) OTHER INFORMATION: /note= "Phosphorylation site"
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 117
(D) OTHER INFORMATION: /note= "Potential N-glycosylation site"
- (ix) FEATURE:
(A) NAME/KEY: Active-site
(B) LOCATION: 18..24
(D) OTHER INFORMATION: /note= "Consensus sequence
predicted to be involved in catalysis"
- (ix) FEATURE:
(A) NAME/KEY: Active-site
(B) LOCATION: 275..278
(D) OTHER INFORMATION: /note= "Consensus sequence
predicted to be involved in catalysis"
- (ix) FEATURE:
(A) NAME/KEY: Active-site
(B) LOCATION: 283..289
(D) OTHER INFORMATION: /note= "Consensus sequence
predicted to be involved in catalysis"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Thr Ser Cys His Xaa Ala Glu Xaa Xaa Ile Xaa Lys Val Ala Ile
1 5 10 15

Phe Gly Gly Thr His Gly Asn Glu Leu Thr Gly Val Phe Leu Val Lys
20 25 30

09065807 200102

His	Trp	Leu	Glu	Asn	Xaa	Xaa	Glu	Ile	Gln	Arg	Thr	Gly	Leu	Glu	Val
35							40						45		
Lys	Pro	Phe	Ile	Thr	Asn	Pro	Arg	Ala	Val	Lys	Cys	Thr	Arg	Tyr	
50						55				60					
Ile	Asp	Cys	Asp	Leu	Asn	Arg	Xaa	Phe	Asp	Xaa	Glu	Asn	Leu	Gly	Lys
65					70					75			80		
Lys	Xaa	Ser	Glu	Asp	Leu	Pro	Tyr	Glu	Val	Arg	Arg	Ala	Gln	Glu	Ile
				85				90					95		
Asn	His	Leu	Phe	Gly	Pro	Lys	Asp	Ser	Glu	Asp	Ser	Tyr	Asp	Ile	Ile
			100				105						110		
Phe	Asp	Leu	His	Asn	Thr	Thr	Ser	Asn	Met	Gly	Cys	Thr	Leu	Ile	Leu
			115				120					125			
Glu	Asp	Ser	Arg	Asn	Xaa	Phe	Leu	Ile	Gln	Met	Phe	His	Tyr	Ile	Lys
			130				135					140			
Thr	Ser	Leu	Ala	Pro	Leu	Pro	Cys	Tyr	Val	Tyr	Leu	Ile	Glu	His	Pro
145					150					155					160
Ser	Leu	Lys	Tyr	Ala	Thr	Thr	Arg	Ser	Ile	Ala	Lys	Tyr	Pro	Val	Gly
					165				170					175	
Ile	Glu	Val	Gly	Pro	Gln	Pro	Gln	Gly	Val	Leu	Arg	Ala	Asp	Ile	Leu
			180				185						190		
Asp	Gln	Met	Arg	Lys	Met	Ile	Xaa	His	Ala	Leu	Asp	Phe	Ile	His	Xaa
			195				200					205			
Phe	Asn	Glu	Gly	Lys	Glu	Phe	Pro	Pro	Cys	Ala	Ile	Glu	Val	Tyr	Lys
			210			215					220				
Ile	Xaa	Xaa	Lys	Val	Asp	Tyr	Pro	Arg	Xaa	Glu	Xaa	Gly	Glu	Ile	Xaa
			225			230				235					240
Ala	Ile	Ile	His	Pro	Xaa	Leu	Gln	Asp	Gln	Asp	Trp	Lys	Pro	Leu	His
				245				250					255		
Pro	Xaa	Asp	Pro	Xaa	Phe	Leu	Thr	Leu	Asp	Gly	Lys	Thr	Ile	Pro	Leu
			260				265					270			
Gly	Gly	Asp	Xaa	Thr	Val	Tyr	Pro	Val	Phe	Val	Asn	Glu	Ala	Ala	Tyr
			275			280						285			
Tyr	Glu	Lys	Lys	Glu	Ala	Phe	Ala	Lys	Thr	Thr	Lys	Leu	Thr	Leu	Asn
			290				295					300			
Ala	Xaa	Ser	Ile	Arg	Xaa	Xaa	Leu	His							
			305				310								

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 313 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

- DRAFT
SEQUENCE SOURCE
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 83
(D) OTHER INFORMATION: /note= "Phosphorylation site"
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 105
(D) OTHER INFORMATION: /note= "Phosphorylation site"
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 108
(D) OTHER INFORMATION: /note= "Phosphorylation site"
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 146
(D) OTHER INFORMATION: /note= "Phosphorylation site"
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 264
(D) OTHER INFORMATION: /note= "Phosphorylation site"
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 117
(D) OTHER INFORMATION: /note= "Potential N-glycosylation site"
- (ix) FEATURE:
(A) NAME/KEY: Active-site
(B) LOCATION: 18..24
(D) OTHER INFORMATION: /note= "Consensus sequence predicted to be involved in catalysis"
- (ix) FEATURE:
(A) NAME/KEY: Active-site
(B) LOCATION: 275..278
(D) OTHER INFORMATION: /note= "Consensus sequence predicted to be involved in catalysis"
- (ix) FEATURE:
(A) NAME/KEY: Active-site
(B) LOCATION: 283..289
(D) OTHER INFORMATION: /note= "Consensus sequence predicted to be involved in catalysis"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
- Met Thr Ser Cys His Val Ala Glu Asp Pro Ile Lys Lys Val Ala Ile
1 5 10 15
- Phe Gly Gly Thr His Gly Asn Glu Leu Thr Gly Val Phe Leu Val Lys
20 25 30
- His Trp Leu Glu Asn Ser Thr Glu Ile Gln Arg Thr Gly Leu Glu Val
35 40 45
- Lys Pro Phe Ile Thr Asn Pro Arg Ala Val Lys Lys Cys Thr Arg Tyr
50 55 60

25xi

Ile Asp Cys Asp Leu Asn Arg Val Phe Asp Pro Glu Asn Leu Gly Lys
65 70 75 80
Lys Lys Ser Glu Asp Leu Pro Tyr Glu Val Arg Arg Ala Gln Glu Ile
85 90 95
Asn His Leu Phe Gly Pro Lys Asp Ser Glu Asp Ser Tyr Asp Ile Ile
100 105 110
Phe Asp Leu His Asn Thr Thr Ser Asn Met Gly Cys Thr Leu Ile Leu
115 120 125
Glu Asp Ser Arg Asn Asp Phe Leu Ile Gln Met Phe His Tyr Ile Lys
130 135 140
Thr Ser Leu Ala Pro Leu Pro Cys Tyr Val Tyr Leu Ile Glu His Pro
145 150 155 160
Ser Leu Lys Tyr Ala Thr Thr Arg Ser Ile Ala Lys Tyr Pro Val Gly
165 170 175
Ile Glu Val Gly Pro Gln Pro Gln Gly Val Leu Arg Ala Asp Ile Leu
180 185 190
Asp Gln Met Arg Lys Met Ile Gln His Ala Leu Asp Phe Ile His Asn
195 200 205
Phe Asn Glu Gly Lys Glu Phe Pro Pro Cys Ala Ile Glu Val Tyr Lys
210 215 220
Ile Met Arg Lys Val Asp Tyr Pro Arg Asn Glu Ser Gly Glu Ile Ser
225 230 235 240
Ala Ile Ile His Pro Lys Leu Gln Asp Gln Asp Trp Lys Pro Leu His
245 250 255
Pro Glu Asp Pro Val Phe Leu Thr Leu Asp Gly Lys Thr Ile Pro Leu
260 265 270
Gly Gly Asp Gln Thr Val Tyr Pro Val Phe Val Asn Glu Ala Ala Tyr
275 280 285
Tyr Glu Lys Lys Glu Ala Phe Ala Lys Thr Thr Lys Leu Thr Leu Asn
290 295 300
Ala Asn Ser Ile Arg Ser Ser Leu His
305 310

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AACCCTACTC TTAAGGAC

25.

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ix) FEATURE:

 - (A) NAME/KEY: modified_base
 - (B) LOCATION: 1
 - (D) OTHER INFORMATION: /mod base= OTHER
/note= "The M13 universal primer tag is attached to base number 1."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCGGGATGAA AATGGAGAA

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ix) FEATURE:

 - (A) NAME/KEY: modified_base
 - (B) LOCATION: 1
 - (D) OTHER INFORMATION: /mod_base= OTHER
/note= "The M13 reverse primer tag is attached to
base 1."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ACCGTGTAA G ATGTAAGC

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AGGATCAAGA CTGGAAACC

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTAAGACACC GTGTAAGATG

15

18

19

20

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Gly Gly Thr His Gly Asn Glu
1 5

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Val Asn Glu Ala Ala Tyr Tyr
1 5

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Val Xaa Glu Xaa Xaa Xaa Tyr
1 5

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

T0T000T-20859560

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Leu Glu Asn Ser Thr Glu Ile Gln Arg Thr Gly Leu Glu Val Lys Pro
 1 5 10 15
 Phe Ile Thr Asn Pro Arg Ala Val Lys Lys
 20 25

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Lys Pro Leu Ile Pro Xaa Asp Pro Val Phe Leu Thr Leu Asp Gly Lys
 1 5 10 15
 Thr Ile Ser Leu Gly Gly Asp Gln Thr Xaa Tyr Pro Xaa Phe Xaa Asn
 20 25 30
 Glu Ala Ala Tyr Tyr
 35

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Xaa Lys Val Asp Tyr Pro Arg Asn Glu Ser Gly Glu Ile Ser Ala Ile
 1 5 10 15
 Ile His Pro Lys Leu Gln Asp Gln
 20

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 16 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Xaa Xaa Xaa Ala Leu Asp Phe Ile Xaa Asn Phe Xaa Glu Xaa Lys Glu
1 5 10 14 15

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 15
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 16
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 18
- (D) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AARGTNGAYT AYCCNNGNAA

20

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 11
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 15
- (D) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TGRTCYTGNA NYTTNGGRTG

20

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 16 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCGTGTACCC AGTGTGTT

16

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CTTCTGAATT GCAGAAATCA

20

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GTAAGACACC GTGTAAGATG

20

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ix) FEATURE:
- (A) NAME/KEY: modified_base
 - (B) LOCATION: 6
 - (D) OTHER INFORMATION: /mod_base= i

- (ix) FEATURE:
- (A) NAME/KEY: modified_base
 - (B) LOCATION: 9
 - (D) OTHER INFORMATION: /mod_base= i

- (ix) FEATURE:
- (A) NAME/KEY: modified_base

- (B) LOCATION: 18
- (D) OTHER INFORMATION: /mod_base= i
- (ix) FEATURE:
 - (A) NAME/KEY: modified_base
 - (B) LOCATION: 21
 - (D) OTHER INFORMATION: /mod_base= i
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGRTANACNG TYTGRTCNCC NCC

23

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ix) FEATURE:
 - (A) NAME/KEY: modified_base
 - (B) LOCATION: 3
 - (D) OTHER INFORMATION: /mod_base= i

- (ix) FEATURE:
 - (A) NAME/KEY: modified_base
 - (B) LOCATION: 6
 - (D) OTHER INFORMATION: /mod_base= i

- (ix) FEATURE:
 - (A) NAME/KEY: modified_base
 - (B) LOCATION: 9
 - (D) OTHER INFORMATION: /mod_base= i

- (ix) FEATURE:
 - (A) NAME/KEY: modified_base
 - (B) LOCATION: 12
 - (D) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCNMGNGCNG TNAARAARTG

20

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Asp Cys Thr Val

1

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(v) FRAGMENT TYPE: internal

(ix) FEATURE:

- (A) NAME/KEY: Active-site
- (B) LOCATION: 1..7
- (D) OTHER INFORMATION: /note= "Consensus sequence of catalytic center in esterases"

(ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /note= "Amino acid 5 is glycine or alanine"

(ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /note= "Amino acid 7 is glutamic acid or aspartic acid"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Gly Xaa Xaa His Xaa Xaa Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(v) FRAGMENT TYPE: internal

(ix) FEATURE:

- (A) NAME/KEY: Active-site
- (B) LOCATION: 1..4
- (D) OTHER INFORMATION: /note= "Consensus sequence of catalytic center in esterases"

(ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /note= "Amino acid 4 is phenylalanine or valine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Asp Xaa Xaa Xaa
1

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(v) FRAGMENT TYPE: internal

(ix) FEATURE:

- (A) NAME/KEY: Active-site
- (B) LOCATION: 1..7
- (D) OTHER INFORMATION: /note= "Consensus sequence of catalytic center in esterases"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Val Xaa Glu Xaa Xaa Xaa Tyr
1 5